A Min-Plus System Theory for Constrained Traffic Regulation and Dynamic Service Guarantees

Cheng-Shang Chang*, Rene L. Cruz †, Jean-Yves Le Boudec‡, and Patrick Thiran §

May 6, 1999

Abstract: By extending the system theory under the $(\min, +)$-algebra to the time varying setting, we solve the problem of constrained traffic regulation and develop a calculus for dynamic service guarantees. For a constrained traffic regulation problem with maximum tolerable delay d and maximum buffer size q, the optimal regulator that generates the output traffic conforming to a subadditive envelope f and minimizes the number of discarded packets is a concatenation of the g-clipper with $g(t) = \min[f(t+d), f(t)+q]$ and the maximal f-regulator. The g-clipper is a *bufferless* device which optimally *drops* packets as necessary in order that its output be conformant to an envelope g. The maximal f-regulator is a *buffered* device that *delays* packets as necessary in order that its output be conformant to an envelope f. The maximal f-regulator is a linear time invariant filter with impulse response f, under the $(\min, +)$-algebra.

To provide dynamic service guarantees in a network, we develop the concept of a dynamic server as a basic network element. Dynamic servers can be joined by concatenation, "filter bank summation," and feedback to form a composite dynamic server. We also show that dynamic service guarantees for multiple input streams sharing a work conserving link can be achieved by a dynamic SCED (Service Curve Earliest Deadline) scheduling algorithm, if an appropriate admission control is enforced.

Keywords: constrained traffic regulation, traffic clipping, service guarantees, min-plus algebra, scheduling, admission control

*Dept. of Electrical Engineering, National Tsing Hua University, Hsinchu 30043 Taiwan, R.O.C. Email: cschang@ee.nthu.edu.tw. This work is supported in part by the National Science Council, Taiwan, R.O.C., under contracts NSC-87-2213-E007-084 and NSC-88-2213-E007-004.

†Dept. of Electrical & Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407. Email: cruz@ece.ucsd.edu. This work was supported in part by the U.S. National Science Foundation under grants NCR 94-15684 and NCR 97-25568, and by the Center for Wireless Communications at UCSD.

‡Institute for Computer Communication and Applications (ICA), Ecole Polytechnique Fédérale de Lausanne (EPFL). `http://icawww.epfl.ch`. This work is supported in part by the Priority Program in Information and Communication Structures of the Swiss National Science, under grant SPP ICS 5003–45311.

§Institute for Computer Communication and Applications (ICA), Ecole Polytechnique Fédérale de Lausanne (EPFL). `http://icawww.epfl.ch`

1 Introduction

Future high speed digital networks aim to provide integrated services, including voice, video, fax, and data. To control interaction among traffic generated by different sources, traffic regulation seems inevitable. In [10], Cruz proposed the following deterministic traffic characterization. A traffic stream, described by a non-decreasing sequence $A \equiv \{A(t), t = 0, 1, 2, \}$ (with $A(0) = 0$), conforms to a function f, called an *envelope*, if

$$A(t) - A(s) \leq f(t - s), \quad \forall s \leq t.$$

Without loss of generality, an envelope f can be assumed to be subadditive [6], i.e., $f(s) + f(t - s) \geq f(t)$ for all $s \leq t$. Using this characterization, a calculus is developed in [10, 11] to compute deterministic performance measures, such as bounds on delay and bounds on queue length. Traffic regulation addresses the problem of modifying a traffic stream so that it conforms to a subadditive envelope f. The problem of traffic regulation was treated systematically in [8, 20]. There it is shown that the optimal traffic regulator that generates an output B conforming to a subadditive envelope f for an input A is a linear time invariant filter with the impulse response f under the $(\min, +)$-algebra, i.e.,

$$B(t) = \min_{0 \leq s \leq t} [A(s) + f(t - s)].$$

We call such a filter the maximal f-regulator. This characterization was also observed in [1][2][27].

As the buffer in the maximal f-regulator is assumed to be infinite, packets from the input might be queued at the regulator. For a real-time service, the delay of a queued packet at the regulator might exceed a maximum tolerable delay and such a packet should be discarded (i.e. clipped). The problem of traffic regulation with such a delay constraint is called the constrained traffic regulation problem in [19]. Its objective is to find a regulator that not only generates traffic conforming to an envelope, but also minimizes the number of discarded packets. In addition to the delay constraint, Konstantopoulos and Ananthram [19] also considered the buffer constraint for the regulator. For $f(t) = \rho t + \sigma$, they derived optimal traffic regulators that satisfied either the delay constraint or the buffer constraint.

Cruz and Taneja [16] considered the zero delay case of the constrained traffic regulation problem. This is also the case without any buffer. By extending the time invariant filtering theory under the $(\min, +)$-algebra to the time varying setting, it is shown there that the departure process of the optimal zero-delay regulator, that generates a departure process conformant to f, is the subadditive closure [8] of the arrival process convolved with f. Such a bufferless regulator is called the f-clipper in [16].

Motivated by all these works, one of the main objectives of this paper is to provide an optimal and implementable solution for the general constrained traffic regulation problem with both the delay constraint and the buffer constraint. As in [16], our approach is based on the time varying filtering theory under the $(\min, +)$-algebra. By extending the subadditive closure in [8] to the time varying setting, we show that the f-clipper with input A and output B can be implemented using the following recursive equation:

$$B(t) = \min\Big[B(t-1) + A(t) - A(t-1), \min_{0 \le s < t}[B(s) + f(t-s)]\Big].$$

The computation complexity of the f-clipper is almost the same as that of the maximal f-regulator. The recursive equation also implies that the f-clipper is greedy. Packets are discarded only when needed.

For the constrained traffic regulation problem with maximum tolerable delay d and maximum buffer size q, the optimal traffic regulator is shown to be a concatenation of the g-clipper with $g(t) = \min[f(t)+q, f(t+d)]$ and the maximal f-regulator. The solution is intuitive as the output from the g-clipper conforms to the envelope g that yields bounded delay d and bounded queue length q at the maximal f-regulator. For example, when $f(t) = \min_{1 \le i \le K}[\rho_i t + \sigma_i]$, the corresponding g-clipper can be implemented by K parallel bufferless $(\sigma_i + \min[q, \rho_i d], \rho_i)$-leaky buckets. A packet is discarded if it cannot be admitted to one of these K leaky buckets. The output from the g-clipper is then fed into K parallel (σ_i, ρ_i)-leaky buckets.

The time varying filtering theory can also be used for dynamic service guarantees. By extending the concept of the service curve in [12, 1, 20] to a bivariate function $F(\cdot, \cdot)$, we define a dynamic F-server for an input A if its output B satisfies

$$B(t) \ge \min_{0 \le s \le t}[A(s) + F(s,t)], \quad \forall t.$$

Analogous to the time invariant filtering theory in [8, 1, 20], a dynamic F-server can be viewed as a linear filter with the time varying impulse response F. It can be combined by concatenation, "filter bank summation," and feedback to form a composite dynamic server. We illustrate the use of the dynamic server by considering a work conserving link with a time varying capacity and a dynamic window flow control problem. We also show that dynamic service guarantees for multiple input streams sharing a work conserving link can be achieved by a dynamic SCED (Service Curve Earliest Deadline) scheduling algorithm, if an appropriate admission control is enforced. As the SCED algorithm in [28], the dynamic SCED algorithm is an EDF (Earliest Deadline First) policy that schedules packets according to their deadlines.

The remainder of the paper is organized as follows. In the next section, we introduce the time varying filtering theory under the $(\min,+)$-algebra. The development is parallel to the time invariant filtering theory in [8, 1, 20]. The reader is also referred to [3], [4], which contains results overlapping with the paper. In Section 3 and Section 4, we introduce the maximal dynamic traffic regulators and the maximal dynamic clippers, respectively. These are used for solving the problem of constrained traffic regulation in Section 5. In Section 6, we develop the concept of dynamic servers and their associated calculus. We show in Section 7 that the dynamic SCED algorithm can be used to achieve dynamic service guarantees. We conclude the paper in Section 8 by discussing possible extensions and applications.

2 Time varying filtering theory under the min-plus algebra

In the section, we introduce the time varying filtering theory under the $(\min,+)$-algebra. The development is parallel to the time invariant filtering theory in [8, 1, 20]. To extend the $(\min,+)$-algebra to the time varying setting, we consider the family of bivariate functions.

$$\tilde{\mathcal{F}} = \{F(\cdot,\cdot) : F(s,t) \geq 0, F(s,t) \leq F(s,t+1), \text{ for all } 0 \leq s \leq t\}$$

Thus, for any $F \in \tilde{\mathcal{F}}$, $F(s,t)$ is nonnegative and non-decreasing in t. For any two bivariate functions F and G in $\tilde{\mathcal{F}}$, we say $F = G$ (resp. $F \leq G$) if $F(s,t) = G(s,t)$ (resp. $F(s,t) \leq G(s,t)$) for all $0 \leq s \leq t$. We define the following two operations for functions in $\tilde{\mathcal{F}}$.

(i) (min) the pointwise minimum of two functions:

$$(F \oplus G)(s,t) = \min[F(s,t), G(s,t)].$$

(ii) (convolution) the convolution of two functions under the $(\min,+)$-algebra:

$$(F \star G)(s,t) = \min_{s \leq \tau \leq t} [F(s,\tau) + G(\tau,t)].$$

One can easily verify that $(\tilde{\mathcal{F}}, \oplus, \star)$ is a complete dioid (see e.g., [5]) with the zero function $\tilde{\epsilon}$ and the identity function $\tilde{\mathbf{e}}$, where $\tilde{\epsilon}(s,t) = \infty$ for all $s \leq t$, and $\tilde{\mathbf{e}}(s,t) = 0$ if $s = t$ and ∞ otherwise. To be precise, we have the following properties.

1. (Associativity) $\forall F, G, H \in \tilde{\mathcal{F}}$,

$$(F \oplus G) \oplus H = F \oplus (G \oplus H),$$
$$(F \star G) \star H = F \star (G \star H).$$

2. (Commutativity) $\forall F, G \in \tilde{\mathcal{F}}$,

$$F \oplus G = G \oplus F.$$

3. (Distributivity for infinite "sums") For any two sequences of functions F_m and G_m in $\tilde{\mathcal{F}}$,

$$(F_1 \oplus F_2 \oplus \ldots \oplus F_m \oplus \ldots) \star (G_1 \oplus G_2 \oplus \ldots \oplus G_m \oplus \ldots)$$
$$= (F_1 \star G_1) \oplus (F_1 \star G_2) \oplus (F_2 \star G_1) \oplus \ldots \oplus (F_m \star G_m) \oplus \ldots$$

4. (Zero element) $\forall F \in \tilde{\mathcal{F}}$,

$$F \oplus \tilde{\epsilon} = F.$$

5. (Absorbing zero element) $\forall F \in \tilde{\mathcal{F}}$,

$$F \star \tilde{\epsilon} = \tilde{\epsilon} \star F = \tilde{\epsilon}.$$

6. (Identity element) $\forall F \in \tilde{\mathcal{F}}$,

$$F \star \tilde{\mathbf{e}} = \tilde{\mathbf{e}} \star F = F.$$

7. (Idempotency of addition) $\forall F \in \tilde{\mathcal{F}}$,

$$F \oplus F = F.$$

The key difference to the time invariant filtering theory is that we do not have the commutative property for $\star$ in $(\tilde{\mathcal{F}}, \oplus, \star)$, i.e., $F \star G \neq G \star H$ in general.

Let $\tilde{\mathcal{F}}_0 = \{F \in \tilde{\mathcal{F}} : F \oplus \tilde{\mathbf{e}} = F\}$. That is, a function $F \in \tilde{\mathcal{F}}_0$ if $F(t,t) = 0$ for all t. As in the time invariant case, we still have the following monotonicity.

8. (Monotonicity) $\forall F \leq \tilde{F}, G \leq \tilde{G}$,

$$F \oplus G \leq \tilde{F} \oplus \tilde{G} \leq \tilde{F},$$
$$F \star G \leq \tilde{F} \star \tilde{G}.$$

If F (resp. G) is in $\tilde{\mathcal{F}}_0$, then $F \star G \leq G$ (resp. $F \star G \leq F$). If both F and G are in $\tilde{\mathcal{F}}_0$, then $F \oplus G \geq F \star G$.

For any function $F \in \tilde{\mathcal{F}}$, define the unitary operator (called the closure operation in this paper)

$$F^* = \lim_{n\to\infty}(F \oplus \tilde{\mathbf{e}})^{(n)} = \lim_{n\to\infty}(\tilde{\mathbf{e}} \oplus F \oplus F^{(2)} \oplus \ldots \oplus F^{(n)}), \tag{1}$$

where $F^{(n)}$ is the self convolution of F for n times, i.e., $F^{(n)} = F^{(n-1)} \star F$, $n \geq 2$ and $F^{(1)} = F$. Expanding (1) yields

$$F^*(s,t) = \inf_S \sum_{i=1}^{m} [F(t_{i-1}, t_i)], \tag{2}$$

where $S = \{t_0, t_1, t_2, \ldots, t_m\}$ is any subset of $\{1, 2, \ldots, t\}$ with $t_0 = s < t_1 < t_2 < \ldots < t_m = t$.

In addition to the algebraic properties, we present several important properties in Lemmas 2.1 and 2.2 that will be used to prove results for constrained traffic regulation and service guarantees.

Lemma 2.1 *Suppose that $F, G \in \tilde{\mathcal{F}}$.*

(i) *(Monotonicity) If $F \leq G$, then $F^* \leq G^*$.*

(ii) *(Closure properties) $F^* = F^* \oplus \tilde{\mathbf{e}} = F^* \star F^* = (F^*)^{(m)} = (F^*)^* \leq F \oplus \tilde{\mathbf{e}} \leq F$.*

(iii) *(Maximum solution) F^* is the maximum solution of the equation $H = (H \star F) \oplus \tilde{\mathbf{e}}$, i.e., for any H satisfying $H = (H \star F) \oplus \tilde{\mathbf{e}}$, $H \leq F^*$.*

(iv) *F^* can be computed recursively from the following equations:*

$$F^*(s,s) = 0,$$

$$F^*(s,t) = \min_{s \leq \tau < t} [F^*(s,\tau) + F(\tau,t)].$$

(v) *$(F \oplus G)^* = (F^* \oplus G^*)^* = (F^* \star G^*)^*$.*

Proof. As the proofs for (i)-(iv) are identical to those in [9, 8], we only prove (v). From the monotonicity, $F \oplus G \geq F^* \oplus G^* \geq F^* \star G^*$. Thus, $(F \oplus G)^* \geq (F^* \star G^*)^*$. On the other hand, one has $F \geq F \oplus G$. Thus, $F^* \geq (F \oplus G)^*$. Similarly, $G^* \geq (F \oplus G)^*$. This implies

$$F^* \star G^* \geq (F \oplus G)^* \star (F \oplus G)^* = (F \oplus G)^*.$$

Thus,

$$(F^* \star G^*)^* \geq ((F \oplus G)^*)^* = (F \oplus G)^*.$$

■

Lemma 2.2 *(Feedback) Suppose that $F, G, H \in \tilde{\mathcal{F}}$.*

(i) *For the equation*

$$H = (H \star F) \oplus G, \tag{3}$$

$H = G \star F^*$ is the maximum solution.

(ii) *If $\inf_t F(t,t) > 0$, then $H = G \star F^*$ is the unique solution.*

(iii) *Under the condition in (ii), if*

$$H \geq (H \star F) \oplus G,$$

then $H \geq G \star F^*$.

The proofs for Lemma 2.2 are identical to those in [9, 8] and thus omitted.

Remark 2.3 As in [8], let $\mathcal{F} = \{f : f(0) \geq 0, f(s) \leq f(t), s \leq t\}$ be the set of nonnegative and non-decreasing functions. Also, let $\mathcal{F}_0$ be the subset of functions in $\mathcal{F}$ with $f(0) = 0$. Then one may define the convolution of a function $f \in \mathcal{F}$ and a bivariate function $G \in \tilde{\mathcal{F}}$ as follows:

$$(f \star G)(t) = \min_{0 \leq s \leq t} [f(s) + G(s,t)].$$

Under such a definition, $f \star G$ is in $\mathcal{F}$. One may view $f \star G$ as a special case of $F \star G$ for some $F \in \tilde{\mathcal{F}}$ with $F(0,t) = f(t)$ for all t and $F(s,t) = \infty$, for all t and $s > 0$. Thus, the results in Lemma 2.2 still hold.

Remark 2.4 A bivariate function F is *time-invariant* if

$$F(s,t) = F(s+u, t+u), \quad \forall s \leq t, \text{ and } u \geq 0.$$

By letting $f(t) = F(0,t)$, one can easily verify that F is time-invariant if and only if there exists some $f \in \mathcal{F}$ such that $F(s,t) = f(t-s)$. As a result, time-invariant bivariate functions commute. To see this, consider two invariant functions F and G and let $f(t) = F(0,t)$ and $g(t) = G(0,t)$. Then

$$(F \star G)(s,t) = \min_{s \leq u \leq t} [f(u-s) + g(t-u)] = (G \star F)(s,t). \tag{4}$$

An important corollary of (4) is that Lemma 2.1(v) can be simplified as follows (cf.[8], Lemma 2.2(xi)):

$$(F \oplus G)^* = (F^* \oplus G^*)^* = (F^* \star G^*)^* = F^* \star G^*. \tag{5}$$

Remark 2.5 A bivariate function F is *additive* if

$$F(s,u)+F(u,t)=F(s,t), \quad \forall s \le u \le t.$$

For an additive bivariate function F, one easily check that

$$F^{(2)}(s,t)=\min_{s\le u\le t}[F(s,u)+F(u,t)]=F(s,t),$$

which implies that $F^* = F$. Note that a bivariate function F is additive if and only if there is a function $f \in \mathcal{F}$ such that $F(s,t)=f(t)-f(s)$. This can be easily verified by choosing $f(t)=F(0,t)$.

3 Dynamic traffic regulation

Given a sequence $A \in \mathcal{F}_0$, it is defined in [10, 11] that A conforms to the (static) upper envelope $f \in \mathcal{F}_0$ if $A(t)-A(s) \le f(t-s)$ for all $s \le t$. It is also shown in [8, 1] that the optimal traffic regulator that generates output traffic conforming to a subadditive envelope f is a linear time invariant filter with the impulse response f under the $(\min,+)$-algebra. In this section, we extend such a result to the time varying setting.

We start from extending the definition of a static envelope to a dynamic envelope.

Definition 3.1 *A sequence $A \in \mathcal{F}_0$ is said to conform to the dynamic upper envelope $F \in \tilde{\mathcal{F}}_0$ if for all $s \le t$ there holds $A(t)-A(s) \le F(s,t)$.*

As in [8, 9], this characterization has the following equivalent statements. The proof is omitted.

Lemma 3.2 *Suppose that $A \in \mathcal{F}_0$ and $F \in \tilde{\mathcal{F}}_0$. The following statements are equivalent.*

(i) *A conforms to the dynamic upper envelope F.*

(ii) $A = A \star F$.

(iii) $A = A \star F^*$.

(iv) *A conforms to the dynamic upper envelope F^*.*

Given a dynamic upper envelope $F \in \tilde{\mathcal{F}}_0$, one can construct a regulator such that for any input $A \in \mathcal{F}_0$, the output from the regulator conforms to the dynamic upper envelope F. This is done in the following theorem. Once again, the proof is omitted.

Theorem 3.3 *Suppose that $A \in \mathcal{F}_0$ and $F \in \tilde{\mathcal{F}}_0$. Let $B = A \star F^*$.*

(i) *(Traffic regulation) B conforms to the dynamic upper envelope F^* and thus B also conforms to the dynamic upper envelope F.*

(ii) *(Flow constraint) $B \leq A$.*

(iii) *(Optimality) For any $\tilde{B} \in \mathcal{F}_0$ that satisfies (i) and (ii), one has $\tilde{B} \leq B$.*

(iv) *(Conformity) A conforms to the dynamic upper envelope F if and only if $B = A$.*

The construction $B = A \star F^*$ is called the maximal dynamic F-regulator *(for the input A).*

As in the time invariant case, the flow constraint $B \leq A$ corresponds to one of the causal conditions in [19] as the number of departures cannot be larger than the number of arrivals. Theorem 3.3(iii) shows that under the flow constraint and the constraint that the output traffic conforms to the dynamic upper envelope F, the maximal F-regulator is the best construction that one can implement.

Example 3.4 (Work conserving link with a time varying capacity) Consider a work conserving link with a time varying capacity. Let $c(t)$ be the maximum number of packets that can be served at time t, $C(t) = \sum_{\tau=1}^{t} c(\tau)$ be the cumulative capacity in the interval $[1, t]$, and $\hat{C}(s,t) = C(t) - C(s)$ be the cumulative capacity in the interval $[s+1, t]$. Let $A(t)$ and $B(t)$ be the input and the output from the work conserving link. Denote by $q(t)$ the number of packets at the link at time t. Then the work conserving link is governed by Lindley's equation

$$q(t+1) = [q(t) + A(t+1) - A(t) - c(t+1)]^+. \tag{6}$$

Suppose $q(0) = 0$. Recursive expansion of Lindley's equation yields

$$q(t) = \max_{0 \leq s \leq t} [A(t) - A(s) - \hat{C}(s,t)]. \tag{7}$$

Since $q(t) = A(t) - B(t)$, we have

$$B(t) = \min_{0 \leq s \leq t} [A(s) + \hat{C}(s,t)].$$

As $\hat{C}$ is an additive bivariate function, we have from Remark 2.5 that $\hat{C}^* = \hat{C}$, which shows that the work conserving link is the maximal dynamic $\hat{C}$-regulator.

We note that a work conserving link with a time varying capacity is also equivalent to a time-varying (greedy) shaper in [20].

Example 3.5 (Traffic regulation with a capacity constraint) Consider a link with a time varying capacity. The link is not necessarily work conserving. As in the previous example, let $c(t)$ be the maximum number of packets that can be served at time t, and $C(t) = \sum_{\tau=1}^{t} c(\tau)$ be the cumulative capacity by time t. Let $A(t)$ and $B(t)$ be the input and the output from the link. Though the link may not be work conserving, the output B is still constrained by the capacity, i.e.,

$$B(t) - B(s) \leq C(t) - C(s). \tag{8}$$

Suppose that we would like to perform traffic regulation for the input A such that the output B conforms to the static envelope $f \in \mathcal{F}$, i.e.,

$$B(t) - B(s) \leq f(t-s), \quad \forall s \leq t. \tag{9}$$

From Theorem 3.3, we know that the optimal implementation for the output to satisfy (8) and (9) is the maximal dynamic F-regulator with

$$F(s,t) = \min[C(t) - C(s), f(t-s)]. \tag{10}$$

If $c(t)$ is bounded above by $c_{\max} > 0$ and if the cumulative time-varying capacity C is bounded below by some curve $h \in \mathcal{F}$ over any time window, that is, if for all $0 \leq s \leq t$, $h(t-s) \leq C(t) - C(s) \leq c_{\max}(t-s)$, then one can derive static service curves bounding below the maximal dynamic F-regulator (10). Such curves are obtained in [15, 21, 23].

4 Dynamic traffic clipping

The maximal dynamic F-regulator solves the traffic regulation problem with infinite buffer. In this section, we consider the traffic regulation problem without buffer. The question is then how one drops packets *optimally* such that the output conforms to a dynamic envelope F. Such a problem was previously solved in [16]; however, the solution in [16] cannot be easily implemented directly. In the following theorem, we present a recursive construction for the solution.

Theorem 4.1 *Suppose that $A \in \mathcal{F}_0$ and $F \in \tilde{\mathcal{F}}_0$. Let $B(t) = (\hat{A} \oplus F)^*(0,t)$, where $\hat{A}(s,t) = A(t) - A(s)$. Then the following statements hold.*

(i) *(Traffic regulation) B conforms to the dynamic upper envelope F.*

(ii) *(Clipping constraint)* $B(t) - B(t-1) \leq A(t) - A(t-1)$ *for all* t.

(iii) *(Optimality) For any* $\tilde{B} \in \mathcal{F}_0$ *that satisfies (i) and (ii), one has* $\tilde{B} \leq B$.

(iv) B *can be constructed by the following recursive equation:*

$$B(t) = \min\Big[B(t-1) + A(t) - A(t-1), \min_{0 \leq s < t}[B(s) + F(s,t)]\Big], \tag{11}$$

with $B(0) = 0$.

(v) *(Conformity)* A *conforms to the dynamic upper envelope* F *if and only if* $B = A$.

The construction in (11) is called the maximal dynamic F-clipper *(for the input A) in the paper.*

Proof. For any $s \leq t$ we have

$$\begin{aligned} B(t) &= (\hat{A} \oplus F)^*(0,t) \\ &\leq (\hat{A} \oplus F)^*(0,s) + (\hat{A} \oplus F)(s,t) \\ &\leq B(s) + F(s,t) \ , \end{aligned}$$

and hence $B \leq B \star F$, so that B is conformant to F, establishing (i).

To see (ii), note similarly that

$$\begin{aligned} B(t) &= (\hat{A} \oplus F)^*(0,t) \\ &\leq (\hat{A} \oplus F)^*(0,t-1) + (\hat{A} \oplus F)(t-1,t) \\ &\leq B(t-1) + \hat{A}(t-1,t) \ . \end{aligned}$$

Next, we establish (iii). Suppose that $\tilde{B} \in \mathcal{F}_0$ satisfies (i) and (ii). Since $\tilde{B}(0) = 0$,

$$\tilde{B} \leq \mathbf{e}, \tag{12}$$

where $\mathbf{e}(0) = 0$ and $\mathbf{e}(s) = \infty$ for $s > 0$. As $\tilde{B}$ conforms to the dynamic envelope F,

$$\tilde{B} \leq \tilde{B} \star F. \tag{13}$$

The inequality in the clipping constraint in (ii) is equivalent to $\tilde{B}(t) - \tilde{B}(s) \leq A(t) - A(s)$ for all $s \leq t$ and it can be rewritten as

$$\tilde{B} \leq \tilde{B} \star \hat{A}, \tag{14}$$

with $\hat{A}(s,t) = A(t) - A(s)$. The constraints in (12)-(14) are equivalent to

$$\tilde{B} = \tilde{B} \oplus (\tilde{B} \star F) \oplus (\tilde{B} \star \hat{A}) \oplus \mathbf{e}. \tag{15}$$

Applying the distributivity and the fact that $\hat{A} \in \tilde{\mathcal{F}}_0$ yields

$$\begin{aligned}\tilde{B} &= (\tilde{B} \star (\tilde{\mathbf{e}} \oplus \hat{A} \oplus F)) \oplus \mathbf{e} \\ &= (\tilde{B} \star (\hat{A} \oplus F)) \oplus \mathbf{e}.\end{aligned}$$

It then follows from Lemma 2.2(i) that $\mathbf{e} \star (\hat{A} \oplus F)^*$ is the maximum solution of (15). Note that

$$(\mathbf{e} \star (\hat{A} \oplus F)^*)(t) = (\hat{A} \oplus F)^*(0,t) = B(t).$$

Thus, B is the maximum solution that satisfies (i) and (ii).

To see (iv), note from Lemma 2.1(iv) that B can be constructed recursively as follows:

$$\begin{aligned}B(t) &= \min_{0 \le s < t} \Big[B(s) + \min[A(t) - A(s), F(s,t)]\Big] \\ &= \min\Big[\min_{0 \le s < t}[B(s) + A(t) - A(s)], \min_{0 \le s < t}[B(s) + F(s,t)]\Big],\end{aligned} \tag{16}$$

with $B(0) = 0$. Since B satisfies the clipping constraint,

$$\begin{aligned}B(s) + A(t) - A(s) &= B(s) + A(t-1) - A(s) + A(t) - A(t-1) \\ &\ge B(s) + B(t-1) - B(s) + A(t) - A(t-1) = B(t-1) + A(t) - A(t-1).\end{aligned}$$

This implies that

$$\min_{0 \le s < t}[B(s) + A(t) - A(s)] = B(t-1) + A(t) - A(t-1).$$

Thus,

$$B(t) = \min\Big[B(t-1) + A(t) - A(t-1), \min_{0 \le s < t}[B(s) + F(s,t)]\Big]. \tag{17}$$

To prove (v), note that if $B = A$, then it follows from (17) that $A = A \star F$. Thus, A conforms to the dynamic envelope F. On the other hand, if A conforms to the dynamic envelope F, then

$$A(t) - A(s) \le F(s,t).$$

This implies $\hat{A} \oplus F = \hat{A}$. As $\hat{A}^* = \hat{A}$,

$$B(t) = (\hat{A} \oplus F)^*(0,t) = A(t) - A(0) = A(t).$$

■

We note that the original representation in [16] is that $B(t) = (\hat{A} \star F^*)^*(0,t)$. This is equivalent to our result in Theorem 4.1 from Lemma 2.1(v) and $\hat{A}^* = \hat{A}$. As $(\hat{A}\oplus F)^* = (\hat{A}\oplus F^*)^*$ in Lemma 2.1(v), one also has the following equivalent implementation

$$B(t) = \min\Big[B(t-1) + A(t) - A(t-1), \min_{0\le s<t}[B(s) + F^*(s,t)]\Big]. \tag{18}$$

Note that the key difference between Theorem 3.3 and Theorem 4.1 is the clipping constraint. The clipping constraint implies that in any given slot, the packets departing are a subset of the packets arriving in the same slot. Let $\ell(t) = A(t) - A(t-1) - (B(t) - B(t-1))$ be the number of packets clipped at time t. From (11), we have

$$\ell(t) = \max\Big[0, B(t-1) + A(t) - A(t-1) - \min_{0\le s<t}[B(s) + F(s,t)]\Big]. \tag{19}$$

Observe from (19) that packet loss occurs at time t only when at least one of the following inequalities is violated

$$(B(t-1) + A(t) - A(t-1)) - B(s) \le F(s,t), \quad s = 0, 1 \ldots, t-1. \tag{20}$$

When this happens, one then discards packets to the extent so that the above inequalities are all satisfied. Note also that (11) implies that the maximal dynamic F-clipper can be implemented in real-time, since the value of $B(t)$ depends only on $B(s-1)$ and $A(s)$ for $s \le t$.

In the following example, we illustrate how one implements the maximal dynamic F-clipper by a work conserving link with a finite buffer when $F(s,t) = \rho(t-s) + q$ for $s < t$.

Example 4.2 (Work conserving link with a finite buffer) Consider the work conserving link with a time varying capacity in Example 3.4. In addition, we assume that the buffer size of the link is q, i.e., at most q packets can be stored at the link. Packets that arrive at the link and find the buffer full are lost. As in Example 3.4, let $A(t)$ and $B(t)$ be the input and the output from the work conserving link. Denote by $q(t)$ the number of packets at the link at time t.

Then we need to modify Lindley's equation in (6) as follows:

$$q(t+1) = \min\Big[[q(t) + A(t+1) - A(t) - c(t+1)]^+, q\Big].$$

The number of lost packets at time t, denoted by $\ell(t)$, is then $\max[q(t-1)+A(t)-A(t-1)-c(t)-q,0]$. Let A_1 be the effective input to the link, i.e.,

$$A_1(t)-A_1(t-1)=A(t)-A(t-1)-\ell(t).$$

For the effective input A_1, the work conserving link behaves like a work conserving link with an infinite buffer. Thus, we have from (7) that

$$q(t)=\max_{0\le s\le t}[A_1(t)-A_1(s)-\hat{C}(s,t)], \tag{21}$$

assuming $q(0)=0$. This then implies

$$\begin{aligned}
\ell(t) &= \max[q(t-1)+A(t)-A(t-1)-c(t)-q,0] \\
&= \max\Big[0, \max_{0\le s\le t-1}[A_1(t-1)-A_1(s)-\hat{C}(s,t-1)]+A(t)-A(t-1)-c(t)-q\Big] \\
&= \max\Big[0, A_1(t-1)+A(t)-A(t-1)-\min_{0\le s<t}[A_1(s)+\hat{C}(s,t)+q]\Big].
\end{aligned}$$

In view of (19), the effective input A_1 to the work conserving link with a finite buffer is in fact the output of the maximal dynamic F-clipper with $F(s,t)=\hat{C}(s,t)+q$, $s<t$. In particular, when $c(t)=\rho$ for all t, we can implement the maximal dynamic F-clipper with $F(s,t)=\rho(t-s)+q$ by the effective input of a work conserving link with constant capacity ρ and buffer q.

For the maximal dynamic F-clipper with the input A and the output B, let $L(t)=A(t)-B(t)$ be the cumulative losses at the clipper by time t. As $B(t)=(\hat{A}\oplus F)^*(0,t)$ in Theorem 4.1, using (2) yields

$$\begin{aligned}
L(t) &= A(t)-\inf_S\sum_{i=1}^m \min[A(t_i)-A(t_{i-1}),F(t_{i-1},t_i)] \\
&= \sup_S\Big[A(t)-\sum_{i=1}^m \min[A(t_i)-A(t_{i-1}),F(t_{i-1},t_i)]\Big] \\
&= \sup_S\Big[\sum_{i=1}^m (A(t_i)-A(t_{i-1}))-\sum_{i=1}^m \min[A(t_i)-A(t_{i-1}),F(t_{i-1},t_i)]\Big] \\
&= \sup_S\sum_{i=1}^m [A(t_i)-A(t_{i-1})-F(t_{i-1},t_i)]^+,
\end{aligned} \tag{22}$$

where $S=\{t_0,t_1,t_2,\ldots,t_m\}$ is any subset of $\{1,2,\ldots,t\}$ with $t_0=0<t_1<t_2<\ldots<t_m=t$. This was previously shown in [16], Corollary 1. A similar result is also obtained in [21] for both the continuous and discrete time settings.

Example 4.3 (Clippers in tandem) Now we compare the output from the maximal dynamic $F_1 \oplus F_2$-clipper and a concatenation of the maximal dynamic F_1-clipper and the maximal dynamic F_2-clipper. Let A be the input to both systems, B_1 be the output from the maximal dynamic F_1-clipper, B_2 be the output from the maximal dynamic F_2-clipper, and B be the output from the maximal dynamic $F_1 \oplus F_2$-clipper. Also let $L(t) = A(t) - B(t)$ be the cumulative losses at the maximal dynamic $F_1 \oplus F_2$-clipper by time t. Similar, let $L_1(t) = A(t) - B_1(t)$ and $L_2(t) = B_1(t) - B_2(t)$. From Theorem 4.1, we have for all $s \leq t$,

$$
\begin{aligned}
B_1(t) - B_1(s) &\leq A(t) - A(s), \\
B_1(t) - B_1(s) &\leq F_1(s, t), \\
B_2(t) - B_2(s) &\leq B_1(t) - B_1(s), \\
B_2(t) - B_2(s) &\leq F_2(s, t).
\end{aligned}
$$

This implies that B_2 conforms to the dynamic upper envelope $F_1 \oplus F_2$ and that $B_2(t) - B_2(t-1) \leq A(t) - A(t-1)$. Thus, $B_2 \leq B$ and $L(t) \leq L_1(t) + L_2(t)$ for all t, by Theorem 4.1. In fact, a concatenation of the maximal dynamic F_1-clipper and the maximal dynamic F_2-clipper is a suboptimal implementation of an $F_1 \oplus F_2$-clipper. The reason for this, as observed in [16], is that the discarding of packets in the F_2-clipper is not accounted for in the F_1-clipper.

Example 4.4 (Clippers in parallel) Continue from the previous example. Since clippers in tandem is suboptimal and may yield more cumulative losses than the optimal one, we may use this to compare the cumulative losses for clippers in parallel. Now suppose both the maximal dynamic F_1-clipper and the maximal dynamic F_2-clipper are fed with the input A. Let B_1' and B_2' be the outputs from these two clippers and $L_1'(t) = A(t) - B_1'(t)$ and $L_2'(t) = A(t) - B_2'(t)$ be the cumulative losses at these two clippers by time t. Clearly, $L_1'(t) = L_1(t)$. It is easy to see from (22) that $L_2'(t) \geq L_2(t)$. Thus, we still have $L(t) \leq L_1'(t) + L_2'(t)$ for all t. This is previously reported in [16], Corollary 2.

5 Constrained traffic regulation

The two traffic regulation problems, with an infinite buffer and without buffer, are two extreme cases. In practice, packets (or cells) may be queued and delayed at a regulator. However, there might be constraints for the buffer size and the delay. In this regard, one might have to discard (i.e. clip) some packets from the input so that the buffer and delay constraints can be satisfied. The question is then how one discard packets *optimally* so that the number of clipped packets can be minimized. Such a problem is called constrained traffic regulation and was first considered in [19] for (σ, ρ)-leaky buckets. Our objective of this section is to provide a general, simple and optimal solution for the constrained traffic regulation problem.

To formalize the problem of constrained traffic regulation with buffer and delay constraints, we let A be the input and B be the output from the regulator. We require that the buffer occupancy in the regulator be less than or equal to q, the delay be bounded above by d, and that the output B be conformant to a dynamic envelope F. Due to these constraints, packets may need to be discarded. Let A_1 be the effective input, i.e. $A_1(t)$ counts the total number of packets arriving up to and including slot t which eventually depart the regulator without being discarded. The objective is to maximize the effective input A_1 and the output B, given the buffer and delay constraints and the constraint that B conforms to the dynamic envelope F. More formally, given the input A and a dynamic envelope F, we seek A_1 and B which are as large as possible subject to the following constraints.

(C1) (Clipping constraint) $A_1(t) - A_1(t-1) \leq A(t) - A(t-1)$ for all t.

(C2) (Buffer constraint) $A_1(t) - q \leq B(t)$ for all t, where q is the buffer size at the regulator.

(C3) (Delay constraint) $A_1(t) \leq B(t+d)$ for all t, where d is the maximum tolerable delay at the regulator.

(C4) (Traffic regulation) B conforms to the dynamic upper envelope F.

(C5) (Flow constraint) $B(t) \leq A_1(t)$ for all t.

The clipping constraint implies that the packets in the effective input A_1 is a subset of the packets in A for any time t. We note that the clipping constraint does not imply that packets arriving at time t have to be clipped at time t. In fact, they could be clipped at some time later than t. However, as we will show below that optimal clipping can be greedy and only those packets arriving at time t need to be clipped at time t. Note also that the natural buffer constraint should be $A_1'(t) - B(t) \leq q$, where $A_1'(t)$ is the cumulative number of packets arriving up to time t which have not been discarded at the end of slot t. Our buffer constraint $A_1(t) - B(t) \leq q$ is fact less restrictive as $A_1(t) \leq A_1'(t)$ for all t. However, as the theorem below shows, the optimal value of $A_1(t)$ can be computed without knowledge of $A(s)$ for $s > t$, so that packets which will eventually be discarded in an optimal clipper can in fact be discarded when they arrive. Assuming this is the case, the backlog of packets in the optimal regulator at the end of slot t is $A_1(t) - B(t)$.

Theorem 5.1 *Suppose that $A \in \mathcal{F}_0$ and $F \in \tilde{\mathcal{F}}_0$. Let A_1 be the output from the maximal dynamic G-clipper for the input A, where*

$$G(s,s) = 0, \quad \forall s, \tag{23}$$

$$G(s,t) = \min[F^*(s,t+d), F^*(s,t)+q], \quad \forall s < t. \tag{24}$$

Also, let B be the output from the maximal dynamic F-regulator for the input A_1. Then all the constraints [C1-5] are satisfied. Moreover, for any $\tilde{A}_1, \tilde{B} \in \mathcal{F}_0$ that satisfy [C1-C5], one has $\tilde{A}_1 \leq A_1$ and $\tilde{B} \leq B$.

The construction of A_1 and B, based on a concatenation of the maximal dynamic G-clipper and the maximal dynamic F-regulator, is called the maximal dynamic F-regulator with delay d and buffer q.

Proof. Suppose that A_1 and B are as stated in the theorem. Theorem 4.1 then implies [C1], and also that $A_1(t) \leq (A_1 \star G)(t)$. Conditions [C4-C5] follow from Theorem 3.3 (i) and (ii). To establish [C2], note that

$$\begin{aligned} A_1(t) - q &\leq (A_1 \star G)(t) - q \\ &\leq \min_{0 \leq s \leq t} [A_1(s) + F^*(s,t) + q] - q \\ &= (A_1 \star F^*)(t) \\ &= B(t) \ . \end{aligned}$$

Similarly, to establish [C3], note that

$$\begin{aligned} A_1(t) &\leq (A_1 \star G)(t) \\ &\leq \min_{0 \leq s \leq t} \{A_1(s) + F^*(s,t+d)\} \ . \end{aligned}$$

Since $A_1(t) \leq A_1(s)$ for $s > t$ and F^* is non-negative, it therefore follows that $A_1(t) \leq (A_1 \star F^*)(t+d) = B(t+d)$, which establishes [C3]. Thus, [C1-C5] are satisfied as claimed.

Next, suppose that $\tilde{A}_1, \tilde{B} \in \mathcal{F}_0$ satisfy [C1-C5]. From Theorem 3.3 (iii), we know that under the flow constraint in [C5] and the traffic constraint [C4], we have

$$\tilde{B} \leq \tilde{A}_1 \star F^*. \tag{25}$$

Moreover, combining this with [C2] and [C3], we obtain

(C2′) (Buffer constraint) $\tilde{A}_1(t) - q \leq (\tilde{A}_1 \star F^*)(t)$ for all t.

(C3′) (Delay constraint) $\tilde{A}_1(t) \leq (\tilde{A}_1 \star F^*)(t+d)$ for all t.

The buffer constraint in (C2′) can be rewritten as

$$\tilde{A}_1 \leq \tilde{A}_1 \star F_2 \tag{26}$$

with $F_2(s,t) = F^*(s,t) + q$. Since $\tilde{A}_1(t) \in \mathcal{F}_0$ is non-decreasing in t and $F^*(s,t)$ is nonnegative,

$$\tilde{A}_1(s) + F^*(s,t+d) \geq \tilde{A}_1(t), \quad s = t+1, \ldots, t+d. \tag{27}$$

Thus, the conditions in (27) are redundant and the delay constraint in (C3′) can be rewritten as

$$\tilde{A}_1 \leq \tilde{A}_1 \star F_3 \tag{28}$$

with $F_3(s,t) = F^*(s,t+d)$. Using the idempotency and distributivity, the constraints in (26) and (28) are equivalent to

$$\tilde{A}_1 = \tilde{A}_1 \oplus \tilde{A}_1 \leq (\tilde{A}_1 \star F_2) \oplus (\tilde{A}_1 \star F_3) = \tilde{A}_1 \star (F_2 \oplus F_3) \tag{29}$$

Note that $(F_2 \oplus F_3)(s,t) = G(s,t)$ for all $s < t$, where G is defined in (23). Thus, $\tilde{A}_1$ conforms to the dynamic envelope G. Using Theorem 4.1(iii) and the assumption that $\tilde{A}_1$ satisfies [C1], it therefore follows that $\tilde{A}_1(t) \leq (\hat{A} \oplus G)^*(0,t) = A_1(t)$. From the monotonicity of $\star$, we also have from (25) that

$$\tilde{B} \leq \tilde{A}_1 \star F^* \leq A_1 \star F^* = B.$$

■

We note that for the special cases that $d = \infty$ (without delay constraint) and that $q = \infty$ (without buffer constraint), the results were previous obtained in Examples 4 and 5 of [21]. The result in Theorem 5.1 not only finds a representation of the optimal traffic regulator that satisfies both the delay constraint and the buffer constraint, but also provides a method for the implementation of such a regulator. In [19], the buffer constraint and the delay constraint are treated separately and it is shown that the optimal solution can be implemented by the greedy flow controller which discards packets only when needed. As shown in Theorem 5.1, the maximal dynamic F-regulator with delay d and buffer q is still the greedy flow controller as the maximal dynamic G-clipper discards packets only when needed.

Example 5.2 (Work conserving link with a finite buffer) In this example, we show that a work conserving link with a finite buffer solves a traffic regulation problem with a buffer constraint. Consider the work conserving link with a time varying capacity and a finite buffer in Example 4.2. As in Example 3.4 and Example 4.2, let A, A_1 and B be the input, the effective input and the output of the link. As we have shown from Example 4.2 that the effective input A_1 to the link is in fact the output of the maximal dynamic G-clipper with $G(s,t) = \hat{C}(s,t) + q$ and from Example 3.4 that the output B from the link is the output from the maximal dynamic F-regulator with $F(s,t) = \hat{C}(s,t)$, the link is a concatenation of the maximal dynamic G-clipper and the maximal dynamic F-regulator. Thus, we have from Theorem 5.1 that the work conserving link with a finite buffer q is the maximal dynamic F-regulator with buffer q, where $F(s,t) = \hat{C}(s,t)$.

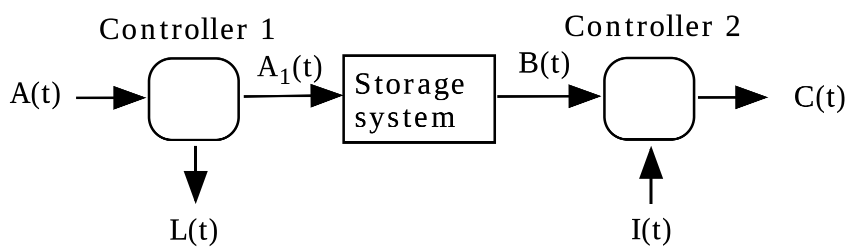

Figure 1: A work conserving link with a finite buffer

There is a well known duality interpretation for a work conserving link with a finite buffer. One may view the cumulative capacity $C(t)$ as the cumulative number of tokens generated by time t. As in a leaky bucket, every packet needs to grab a token for its departure. Thus, packet losses occur when the buffer is full and token losses occur when the buffer is empty. To be precise, let $q(t)$ be the number of packets at the link at time t, $L(t) = A(t) - A_1(t)$ be the cumulative number of packet losses by time t, and $I(t) = C(t) - B(t)$ be the cumulative number of token losses by time t. Figure 1 represents this system. Then one has the following conditions of complementary slackness:

$$\mathbf{1}\{q(t) < q\}(L(t) - L(t-1)) = 0, \quad \text{for all } t,$$
$$\mathbf{1}\{q(t) > 0\}(I(t) - I(t-1)) = 0, \quad \text{for all } t.$$

As

$$q(t) = A_1(t) - B(t) = (A(t) - C(t)) + I(t) - L(t),$$

the work conserving link with a finite buffer solves the so called Skorokhod reflection problem with two boundaries [29], where $A(t) - C(t)$ is the free process, $I(t)$ is the lower boundary process, and $L(t)$ is the upper boundary process (see e.g., [19, 18] for more detailed discussions of the reflection problem). Since a work conserving with a finite buffer also solves the buffer-constrained traffic regulation problem, it follows from (22) that the upper boundary process of the reflection problem admits the following close form representation (in terms of the free process):

$$L(t) = \sup_{S} \sum_{i=1}^{m} [(A(t_i) - C(t_i) - (A(t_{i-1}) - C(t_{i-1})) - q]^+,$$

where $S = \{t_0, t_1, t_2, \ldots, t_m\}$ is any subset of $\{1, 2, \ldots, t\}$ with $t_0 = 0 < t_1 < t_2 < \ldots < t_m = t$. Using (2) and $B = A_1 \star \hat{C}$, one can also show that the lower boundary process admits the following close form representation:

$$I(t) = \sup_{S} \sum_{i=1}^{m-1} \max[(C(t_i) - A(t_i) - (C(t_{i-1}) - A(t_{i-1})), -q].$$

We also note the queue length process $q(t)$ can also be represented in close form. Two representations based on min,max and plus operations were given in [14].

Example 5.3 (Multiple leaky buckets with delay and buffer constraints) Now consider the maximal dynamic F-regulator with delay d and buffer q when

$$F(s, s+t) = \min_{1 \le i \le K} [\rho_i t + \sigma_i], t > 0.$$

This corresponds to the case of multiple leaky buckets with the delay constraint d and the buffer constraint q. In this case,

$$\begin{aligned} G(s,t) &= \min\Big[\min_{1 \le i \le K} [\rho_i(t + d - s) + \sigma_i], \min_{1 \le i \le K} [\rho_i(t - s) + \sigma_i] + q\Big] \\ &= \min_{1 \le i \le K} \Big[\rho_i(t - s) + \sigma_i + \min[q, \rho_i d]\Big]. \end{aligned}$$

Thus, one can construct the maximal dynamic G-clipper by feeding the input to K parallel bufferless $(\sigma_i + \min[q, \rho_i d], \rho_i)$-leaky buckets. A packet is discarded (or clipped) if it cannot be admitted to one of these K leaky buckets. The output from the maximal dynamic G-clipper is then fed into another K parallel (σ_i, ρ_i)-leaky buckets with buffer q.

To bound the cumulative loss for the maximal dynamic G-clipper in this example, we may apply the comparison result in Example 4.4. Consider K maximal dynamic clippers, all subject to the same input. The i^{th} clipper is the maximal dynamic G_i-clipper with

$$G_i(s,t) = \rho_i(t-s) + \sigma_i + \min[q, \rho_i d].$$

Let $L_i(t)$ be the cumulative number of losses by time t at the i^{th} clipper. From Example 4.4, $\sum_{i=1}^{K} L_i(t)$ is an upper bound for the cumulative loss for the maximal dynamic G-clipper. Now $L_i(t)$ is much easier to compute as it is simply the cumulative loss for a work conserving link with capacity ρ_i and buffer $\sigma_i + \min[q, \rho_i d]$ in Example 5.2.

Example 5.4 (Bounding losses by segregation between buffer and policer) We have shown in Theorem 5.1 that the maximal dynamic F'-regulator with buffer q is the optimal implementation of the constrained traffic regulation problem that generates an output that conforms to the dynamic envelope F' subject to the buffer constraint q. In this example, we will show that segregation of buffer discard and policing discard provides an upper bound on the cumulative losses for the maximal dynamic F'-regulator with buffer q.

As we have shown in Theorem 5.1, the first stage of the maximal dynamic F'-regulator with buffer q is the maximal dynamic F-clipper, where

$$F(s,t) = F'(s,t) + q. \tag{30}$$

Let $A(t)$, $D(t)$ and $L(t)$ be its input, output, and the cumulative losses by time t, i.e., $L(t) = A(t) - D(t)$. We now compare the cumulative losses $L(t)$ with the losses in another system made of two parts, as shown in Figure 2. The first part is some causal system with storage capacity q. We know however that the first part discards packets as soon as the total backlogged packets in this system exceeds q. This operation is called *buffer discard*, and the amount of buffer discarded packets by time t is denoted by $L_{\mathrm{Buf}}(t)$. The second part is the maximal dynamic F'-clipper called here policer. Packets are discarded as soon as the total output of the storage system exceeds the maximum output allowed by the policer. This operation is called *policing discard*, and the amount of discarded packets by time t due to policing is denoted by $L_{\mathrm{Pol}}(t)$.

We show that $L(t) \le L_{\mathrm{Buf}}(t) + L_{\mathrm{Pol}}(t)$. Let $B_1(t)$ be the output of the buffer clipper, $A_2(t)$ and $B_2(t)$ be respectively the input and the output of the policer clipper. As B_2 is the output of the maximal dynamic F'-clipper,

$$B_2(t) - B_2(s) \le F'(s,t). \tag{31}$$

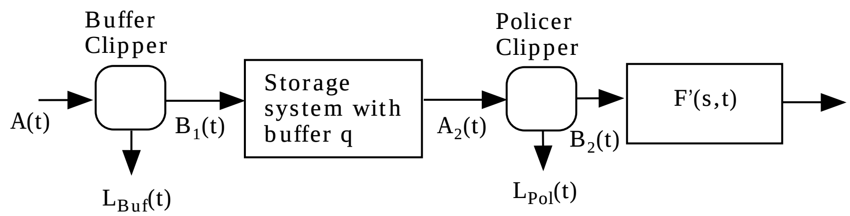

Figure 2: A storage/policer system with separation between losses due to buffer discard and to policing discard

Now let A_1 be the effective input to the system, i.e.,

$$A_1(t) = A(t) - L_{\text{Buf}}(t) - L_{\text{Pol}}(t). \tag{32}$$

Also, as shown in Figure 2, we have

$$L_{\text{Buf}}(t) = A(t) - B_1(t), \tag{33}$$

and

$$L_{\text{Pol}}(t) = A_2(t) - B_2(t). \tag{34}$$

Since $L_{\text{Buf}}(t) + L_{\text{Pol}}(t)$ is a non-decreasing function in t, we have from (32) that

$$A_1(t) - A_1(s) \leq A(t) - A(s). \tag{35}$$

On the other hand, because the "storage system" is causal, it satisfies the flow constraint

$$A_2(t) \leq B_1(t). \tag{36}$$

Since its storage space is limited to q, we also have

$$B_1(t) \leq A_2(t) + q. \tag{37}$$

Using (32) and (33), we have for all $0 \leq s < t$,

$$A_1(t) - A_1(s) = B_1(t) - B_1(s) - (L_{\text{Pol}}(t) - L_{\text{Pol}}(s)).$$

From (36), (37), (34), (31) and (30), it then follows

$$\begin{aligned} A_1(t) - A_1(s) &\leq A_2(t) - A_2(s) - (L_{\text{Pol}}(t) - L_{\text{Pol}}(s)) + q \\ &= B_2(t) - B_2(s) + q \\ &\leq F'(s,t) + q = F(s,t) \end{aligned} \tag{38}$$

Combining (35) with (38), one notices that A_1 satisfies the same constraints as D. As D is the output from the optimal implementation in Theorem 5.1, it follows that $A_1(t) \leq D(t)$, or equivalently that $L(t) \leq L_{\mathrm{Buf}}(t) + L_{\mathrm{Pol}}(t)$.

Such a separation of resources between "buffered system" and "policing system" is used in the estimation of loss probability for devising statistical CAC algorithms as proposed by Lo Presti et al. [26] (see also Elwalid et al [17]).

6 Dynamic service guarantees

To guarantee end-to-end deterministic quality-of-service for an input, the concept of service curves is developed in [12, 1, 20] to work with the static envelopes in Cruz [10, 11]. A server is called a static f-server ($f \in \mathcal{F}_0$) for an input sequence A if its output sequence $B \equiv \{B(t), t = 0, 1, 2, \ldots\}$ satisfies

$$B(t) \geq \min_{0 \leq s \leq t} [A(s) + f(t-s)] \tag{39}$$

for all t. Based on this, there is an associated filtering theory (under the $(\min, +)$-algebra) in [8, 1, 20] that eases design and computation of deterministic QoS. Our main objective of this section is to extend the concept of service curves and the associated filtering theory to the time varying setting so that *dynamic* QoS can be guaranteed. The theory is based on the following definition of a dynamic F-server.

Definition 6.1 *(Dynamic F-server) A server is called a dynamic F-server ($F \in \tilde{\mathcal{F}}_0$) for an input sequence A if its output sequence satisfies $B \geq A \star F$, i.e.,*

$$B(t) \geq \min_{0 \leq s \leq t} [A(s) + F(s,t)] \tag{40}$$

for all t. If the inequality in (40) is satisfied for all input sequences, then we say the dynamic F-server is universal. If the inequality in (40) is an equality, we say the dynamic F-server is exact.

Analogous to the filtering theory for static service curves, one may view the right hand side of (40) as the output from a linear filter with the time varying impulse response $F(s,t)$ under the $(\min, +)$-algebra. If F is a time-invariant bivariate function, then the dynamic F-server is equivalent to a static f-server, where $f(t) = F(0,t)$.

Clearly, the maximal dynamic F-regulator is a *universal* and *exact* dynamic F^*-server. Analogous to the time invariant case, one has the following properties in Theorems 6.2-6.5 for dynamic F-servers. The proofs are omitted as they are identical to those in [8, 9].

Theorem 6.2 *(Concatenation) A concatenation of a dynamic F_1-server for an input sequence A and a dynamic F_2-server for the output from the dynamic F_1-server is a dynamic F-server for A, where $F = F_1 \star F_2$.*

Theorem 6.3 *(Filter bank summation) Consider an input sequence A. Let B_1 (resp. B_2) be the output from a dynamic F_1-server (resp. F_2-server) for A. The output from the "filter bank summation", denoted by B, is $B_1 \oplus B_2$. Then the "filter bank summation" of a dynamic F_1-server for A and a dynamic F_2-server for A is a dynamic F-server for A, where $F = F_1 \oplus F_2$.*

Theorem 6.4 *(Feedback) Consider an input sequence $A \in \mathcal{F}_0$ and a dynamic F-server for B, where $B = A \oplus A_1$, and A_1 is the output from the dynamic F-server. If $\inf_t F(t,t) > 0$, then the feedback system is a dynamic F^*-server for A.*

Theorem 6.5 *Consider a dynamic F_2-server for A. Let B be the output. Also, let $q = \sup_{t \geq 0}[A(t) - B(t)]^+$ be the maximum queue length at the server, where $x^+ = \max(0, x)$. Let $d = \inf\{\delta \geq 0 : B(t+\delta) \geq A(t) \text{ for all } t\}$ be the maximum delay at the server. Suppose that A conforms to the dynamic upper envelope F_1.*

(i) *(Queue length) $q \leq \sup_s \max_{t \geq s}[F_1^*(s,t) - F_2(s,t)]^+$.*

(ii) *(Output burstiness) If $B \leq A$, then B conforms to the dynamic upper envelope F_3^*, where*

$$F_3(s,t) = \max_{0 \leq \tau \leq s}[F_1^*(\tau,t) - F_2(\tau,s)]^+.$$

(iii) *(Delay) $d \leq \inf\{\delta \geq 0 : \sup_s \max_{t \geq s}[F_1^*(s,t) - F_2(s,t+\delta)] \leq 0\}$.*

Remark 6.6 As the maximal dynamic F-regulator is a dynamic F^*-server, there is an intuitive explanation why the maximal dynamic F-regulator with delay d and buffer q is a concatenation of the maximal dynamic G-clipper (with G being defined in (23)) and the maximal dynamic F-regulator. As shown in Theorem 4.1, the output from the maximal dynamic G-clipper conforms to the dynamic envelope $G(s,t) = \min[F^*(s,t+d), F^*(s,t) + q]$. When such an output is fed to the maximal dynamic F-regulator, one has from Theorem 6.5 that the delay at the maximal dynamic F-regulator is bounded above by d and the queue length is also bounded above by q. Thus, both the delay constraint and the buffer constraint are satisfied.

In the following, we illustrate the use of dynamic service guarantees by a dynamic window flow control problem.

Example 6.7 (Dynamic window flow control) Consider a network with the input A and the output B. Suppose that the network enforces a dynamic window flow control for the input A with the dynamic window size $w(t)$. We assume that $\inf_t w(t) > 0$. For the dynamic window flow control system, the effective input to the network, denoted by A_1, satisfies

$$A_1(t) = \min[A(t), B(t) + w(t)]. \tag{41}$$

Observe that $B(t) + w(t) = (B \star H)(t)$, where H is the function with $H(s,t) = \infty$ for $s < t$ and $H(t,t) = w(t)$. One may rewrite (41) as follows:

$$A_1 = A \oplus (B \star H). \tag{42}$$

Also, we assume that the network is a dynamic F-server for the effective input A_1, i.e.,

$$B \geq A_1 \star F. \tag{43}$$

In conjunction with (42),

$$B \geq A_1 \star F = (A \oplus (B \star H)) \star F = (A \star F) \oplus (B \star (H \star F)),$$

where we apply the distributive property and the associativity of $\star$. Since we assume that $\inf_t w(t) > 0$,

$$\inf_t (H \star F)(t,t) = \inf_t [H(t,t) + F(t,t)] \geq \inf_t w(t) > 0.$$

We then have from Lemma 2.2(iii) that

$$B \geq A \star F \star (H \star F)^*.$$

Thus, the dynamic window flow control system is a dynamic $F \star (H \star F)^*$-server.

7 The dynamic SCED scheduling algorithm

In this section, we define a scheduling algorithm, called the dynamic SCED algorithm, which we will show achieves the dynamic service guarantees in Section 6.

Consider a server with a time varying capacity. Let $c(t)$ be the maximum number of packets that can be served at time t, and $\hat{C}(s,t) = \sum_{\tau=s+1}^{t} c(\tau)$ be the cumulative capacity in the interval $[s+1,t]$. A policy is called the Earliest Deadline First (EDF) if the server schedules the packets according to their deadlines. Note that the EDF policy is work conserving, i.e., the server serves packets whenever there are packets at the server.

Now consider feeding n streams of inputs to such a server. Let $A_i(t)$ be the cumulative number of packet arrivals of the i^{th} stream up to time t. Each packet is assigned a deadline. We assume that the deadlines within the same stream are *non-decreasing*. Also, let $N_i(t)$ be the number of packets from the i^{th} stream that have deadlines not greater than t. As we assume the deadlines for each stream is non-decreasing, packet k from stream i is assigned the deadline $D_{i,k}$ from the following inverse mapping

$$D_{i,k} = \inf\{t : t \geq 0 \text{ and } N_i(t) \geq k\} \ . \tag{44}$$

Theorem 7.1 *Suppose that the server is operated under the EDF policy.*

(i) *A necessary condition for every packet to be served not later than its deadline is*

$$\sum_{i=1}^{n} N_i(t) \leq \min_{0\leq s\leq t}[\sum_{i=1}^{n} A_i(s) + \hat{C}(s,t)], \tag{45}$$

for all t.

(ii) *A sufficient condition for every packet to be served not later than its deadline is*

$$\sum_{i\in S} N_i(t) \leq \min_{0\leq s\leq t}[\sum_{i\in S} A_i(s) + \hat{C}(s,t)], \tag{46}$$

for all t and for every S that is a subset of $\{1,2,\ldots,n\}$.

Proof. (i) Let $B(t)$ be the cumulative number of packet departures from all streams up to time t. Since the EDF policy is work conserving, we have from Example 3.4 that

$$B(t) = \min_{0\leq s\leq t}[\sum_{i=1}^{n} A_i(s) + \hat{C}(s,t)].$$

As we assume that every packet is served not later than its deadline, $\sum_{i=1}^{n} N_i(t) \leq B(t)$ for all t.

(ii) We prove this by contradiction as in [25]. Suppose that the first packet that misses its deadline occurs at time t. Let τ^* be the last slot no later than t such that the server serves less than $c(\tau^*)$ packets. Since the EDF policy is work conserving, $\tau^* < t$ as there are at least one stream i packet backlogged at time t. Moreover, there are exactly $\hat{C}(\tau^*, t)$ packets served in the interval $[\tau^* + 1, t]$.

Now let s^* be the last slot in the interval $[\tau^* + 1, t]$ during which a packet with deadline greater than t is served. If all the packets served during the interval $[\tau^* + 1, t]$ have deadlines less than or equal to t, then define $s^* = \tau^*$ (in this case, there are no backlogged packets at the end of slot s^*). Thus, during the interval $[s^* + 1, t]$, exactly $\hat{C}(s^*, t)$ packets are served, and each of these packets has a deadline that is less than or equal to t.

Let S be the set of streams that are not backlogged at the end of slot s^*. We claim that those packets served in $[s^* + 1, t]$ can only come from the streams in S. Suppose that stream i is not in S. Since there is a packet with deadline greater than t is served in slot s^*, all the backlogged stream i packets at the end of slot s^* must have deadlines greater than t. This implies all the stream i packets with deadlines not greater than t have been served as we assume the deadlines are non-decreasing within the same stream. Thus, those packets served in $[s^* + 1, t]$ can only come from the streams in S as those packets have deadlines less than or equal to t.

Now suppose that stream i is in S. As there are no backlogged stream i packets at the end of slot s^*, all the stream i packets that arrive not later than s^* have been served. Thus, the number of stream i packets that can be served in $[s^*+1, t]$ is bounded above by $(N_i(t) - A_i(s^*))^+$. This in turn implies that the number of packets served in $[s^* + 1, t]$ is bounded above by $\sum_{i \in S}(N_i(t) - A_i(s^*))^+$. As there is a packet that misses its deadline at time t, the bound is strict. Thus,

$$\hat{C}(s^*, t) < \sum_{i \in S}(N_i(t) - A_i(s^*))^+ = \sum_{i \in S'}[N_i(t) - A_i(s^*)]$$

for some S' that is a subset of S with $N_i(t) \geq A_i(s^*)$. As S' is a subset of $\{1, 2, \ldots, n\}$, we have a contradiction to (46). ■

Lemma 7.2 *Suppose we choose $N_i = A_i \star F_i$ some $F_i \in \tilde{\mathcal{F}}_0$, $i = 1, \ldots, n$. If $\sum_{i=1}^n F_i(s,t) \leq \hat{C}(s,t)$ for all $0 \leq s \leq t$, then all the packets are served not later than their deadlines.*

Such a deadline assignment scheme is called the dynamic SCED algorithm in this paper.

Proof. It suffices to verify that the sufficient condition in Theorem 7.1 is satisfied. Note that for every S in $\{1, 2, \ldots, n\}$

$$\sum_{i \in S} N_i(t) = \sum_{i \in S} \min_{0 \leq s \leq t}[A_i(s) + F_i(s,t)] \leq \min_{0 \leq s \leq t} \sum_{i \in S}[A_i(s) + F_i(s,t)]$$

$$= \min_{0\le s\le t}[\sum_{i\in S} A_i(s) + \sum_{i\in S} F_i(s,t)] \le \min_{0\le s\le t}[\sum_{i\in S} A_i(s) + \sum_{i=1}^{n} F_i(s,t)] \le \min_{0\le s\le t}[\sum_{i\in S} A_i(s) + \hat{C}(s,t)]$$

where we use $F_i(s,t) \ge 0$ and $\sum_{i=1} F_i(s,t) \le \hat{C}(s,t)$ in the last two inequalities. ■

The next lemma implies that deadlines in the dynamic SCED algorithm can be assigned in real-time. Specifically, if packet k from stream i arrives during slot t, $D_{i,k}$ can be computed without knowledge of $A_i(s)$ for $s > t$.

Lemma 7.3 *Suppose packet k from stream i arrives during slot t. Then under the dynamic SCED algorithm, $D_{i,k} = D_{i,k}(t)$ where*

$$D_{i,k}(t) = \inf\{\Delta : \Delta \ge t \textit{ and } \min_{0\le u\le t-1}[A_i(u) + F_i(u,\Delta)] \ge k\} \ . \tag{47}$$

Proof. Note that under the dynamic SCED algorithm

$$D_{i,k} = \inf\{\Delta : \Delta \ge 0 \text{ and } (A_i \star F_i)(\Delta) \ge k\} \ . \tag{48}$$

Since packet k arrives at time t we have $A_i(u) < k$ for $u < t$. Thus, $(A_i \star F_i)(\Delta) \le A_i(\Delta) < k$ when $\Delta \le t-1$, which implies that $D_{i,k} \ge t$ by definition of $D_{i,k}$ in (48). Therefore, by definition of $D_{i,k}$ we have

$$\begin{aligned} k &\le (A_i \star F_i)(D_{i,k}) \\ &\le \min_{0\le u\le t-1}[A_i(u) + F_i(u, D_{i,k})] \ . \end{aligned}$$

By definition of $D_{i,k}(t)$, this implies $D_{i,k}(t) \le D_{i,k}$. To show the reverse inequality, note that by definition of $D_{i,k}(t)$ we have

$$\min_{0\le u\le t-1}[A_i(u) + F_i(u, D_{i,k}(t))] \ge k \ . \tag{49}$$

Since $A_i(u) \ge k$ for $u \ge t$ and F_i is non-negative, inequality (49) implies that $(A_i \star F_i)(D_{i,k}(t)) \ge k$. By definition of $D_{i,k}$, this then implies that $D_{i,k} \le D_{i,k}(t)$. ■

In Theorem 7.4, we state the admission criteria for the dynamic SCED algorithm for a server with a time varying capacity.

Theorem 7.4 *A set of n arrival streams, indexed $i = 1, \dots, n$, arrives to a server. The arrival sequence of the i^{th} stream is denoted by A_i, and is known to conform to the dynamic upper envelope G_i. The server has a time varying capacity to serve up to $c(t)$ packets during slot t. Under the dynamic SCED algorithm, the server is a dynamic F_i-server for A_i, for all $i = 1, \dots, n$ if the following condition is satisfied for all $s \le t$:*

$$\sum_{i=1}^{n} (G_i \star F_i)(s,t) \le \hat{C}(s,t). \tag{50}$$

Proof. As we assume that A_i conforms to the dynamic upper envelope G_i, we have from Lemma 3.2(ii) that $A_i = A_i \star G_i$. Thus,

$$N_i = A_i \star F_i = (A_i \star G_i) \star F_i = A_i \star (G_i \star F_i),$$

where we apply the associativity of $\star$. From Lemma 7.2, it then follows that all the packets are served before their deadlines. Denote by $B_i(t)$ the cumulative number of departures from stream i by time t. Thus,

$$B_i \ge N_i = A_i \star F_i$$

and the server is a dynamic F_i-server for A_i, for all $i = 1, \dots, n$. ∎

8 Conclusions

By extending the filtering theory under the $(\min, +)$-algebra to the time varying setting, we solved the problem of constrained traffic regulation. For a constrained traffic regulation problem with maximum tolerable delay d and maximum buffer size q, we showed that the optimal regulator that generates the output traffic conforming to a dynamic envelope F and minimizes the number of discarded packets is a concatenation of the maximal dynamic G-clipper with $G(s,t) = \min[F^*(s,t+d), F^*(s,t)+q]$ and the maximal dynamic F-regulator. To provide dynamic service guarantees in a network, we developed the concept of the dynamic F-server as a basic network element. We showed that dynamic servers can be joined by concatenation, "filter bank summation," and feedback to form a composite dynamic server. We also proposed the dynamic SCED scheduling algorithm to achieve dynamic service guarantees for a work conserving link subject to multiple inputs.

One possible application of the time varying filtering theory is dynamic admission control. For a given connection i, we may define a service curve f_i to be guaranteed over the interval

$[a_i+1, b_i]$ if a dynamic service curve F_i is guaranteed, where

$$F_i(s,t) = \begin{cases} 0 & \text{, if } s \le a_i \text{ and } t \le a_i \\ f_i(t-a_i) & \text{, if } s \le a_i \text{ and } a_i \le t \le b_i \\ f_i(t-s) & \text{, if } a_i \le s \le b_i \text{ and } a_i \le t \le b_i \\ f_i(b_i-s) & \text{, if } a_i \le s \le b_i \text{ and } t > b_i \\ 0 & \text{, if } s \ge b_i \text{ and } t \ge b_i \\ f_i(b_i-a_i) & \text{, if } s < a_i \text{ and } t > b_i \text{ .} \end{cases}$$

For such a definition for dynamic service guarantees, an interesting problem is to find the relaxation time r_i such that connection i has virtually no impact on the admission criteria in Theorem 7.4 after $b_i + r_i$.

Finally, we note that our approach is also applicable in the continuous-time setting, as shown in [21, 22, 23]. We also note that the bivariate function F could be *random*. By specifying the probabilistic characteristics of the bivariate function F, it is possible to provide probabilistic guarantees. Previous results along this line could be found in [7, 13].

References

[1] R. Agrawal, R.L. Cruz, C. Okino, and R. Rajan, "Performance bounds for flow control protocols," April, 1998. Available from http://www.ece.wisc.edu/~agrawal.

[2] R. Agrawal and R. Rajan, "A general framework for analyzing schedulers and regulators in integrated services network," *Proceedings of the 34th Annual Allerton Conference on Communication, Control, and Computing*, pp. 239-248, Oct. 1996.

[3] R. Agrawal and R. Rajan, "Open and closed loop control in integrated services networks," *Proceeding of the 36th IEEE CDC*, Vol. 2, pp. 1798-1803, Dec. 1997.

[4] R. Agrawal, F. Baccelli and R. Rajan, "An algebra for queueing networks with time varying service and its application to the analysis of integrated service networks," Technical Report ECE-98-2, ECE Dept., University of Wisconsin-Madison, May 1998. Available from http://www.ece.wisc.edu/~agrawal.

[5] F. Baccelli, G. Cohen, G. Oslder, and J. Quadrat, *Synchronization and Linearity: An Algebra for Discrete Event Systems,* Wiley, 1992.

[6] C.S. Chang, "Stability, queue length and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, Vol. 39, pp. 913-931, 1994.

[7] C.S. Chang, "On the exponentiality of stochastic linear systems under the max-plus algebra," *IEEE Transactions Automatic Control*, Vol. 41, pp. 1182-1188, 1996.

[8] C.S. Chang, "On deterministic traffic regulation and service guarantees: a systematic approach by filtering," *IEEE Transactions on Information Theory*, Vol. 44, pp. 1097-1110, 1998.

[9] C.S. Chang, "Matrix extensions of the filtering theory for deterministic traffic regulation and service guarantees," *IEEE J. Selected Areas in Communications*, Vol. 16, pp. 708-718, 1998.

[10] R.L. Cruz, "A calculus for network delay, Part I: Network elements in isolation," *IEEE Transactions on Information Theory*, Vol. 37, pp. 114-131, 1991.

[11] R.L. Cruz, "A calculus for network delay, Part II: Network analysis," *IEEE Transactions on Information Theory*, Vol. 37, pp. 132-141, 1991.

[12] R.L. Cruz, "Quality of service guarantees in virtual circuit switched networks," *IEEE J. Selected Areas in Communications*, Vol. 13, pp. 1048-1056, 1995.

[13] R.L. Cruz, "Quality of service management in integrated services networks," *Proc. 1st Semi-Annual Research Review*, CWC UCSD, June, 1996.

[14] R.L. Cruz and H.-N. Liu, "Single server queues with loss: a formulation," *Proc. 1993 CISS*, Johns Hopkins Univ., March 1993.

[15] R.L. Cruz and C.M. Okino. 'Service guarantees for a flow control', Preprint, a first version also in 34th Allerton Conf. on Comm., Cont., and Comp. Monticello, IL, Oct. 1996.

[16] R.L. Cruz and M. Taneja, "An analysis of traffic clipping," *Proc. 1998 Conf. on Information Science & Systems*, Princeton University.

[17] A. Elwalid, D. Mitra and R. Wenworth, "A new Approach for Allocating Buffers and Bandwidth to Heterogeneous, Regulated Traffic in ATM node," *IEEE J. Selected Areas in Communications*, Vol. 13(6), pp. 1048–1056, 1995.

[18] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems*, New-York, Wiley, 1985.

[19] T. Konstantopoulos and V. Anantharam. "Optimal flow control schemes that regulate the burstiness of traffic," *IEEE/ACM Transactions on Networking* , Vol. 3, pp. 423-432, 1995.

[20] J.Y. Le Boudec, "Application of network calculus to guarantee service networks," *IEEE Transactions on Information Theory*, Vol. 44, pp. 1087-1096, 1998.

[21] J.Y. Le Boudec and P. Thiran, "Network calculus viewed as a min-plus system theory applied to communication networks," *Technical Report SSC/1998/016.* Available from http://icawww.epfl.ch.

[22] J.Y. Le Boudec and P. Thiran, "A Note on Time and Space Methods in Network calculus," *Proceedings of the International Zurich Seminar on Broadband Communications*, Zurich, Switzerland, pp. 267-272, February 1998.

[23] J.Y. Le Boudec and P. Thiran, "Network Calculus using Min-Plus System Theory" in *High Performance Networks for Multimedia Applications*, Edited by A. Danthine, O. Spaniol, W. Effelsberg and D. Ferrari, Kluwer, 1999.

[24] J.Y. Le Boudec and A. Ziedins, "A CAC algorithm for VBR connections over a VBR trunk," *Proceedings of ITC 15*, pp. 59-70, Washington, June 1997.

[25] C.L Liu and J.W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," *Journal of the Association for computing Machinery*, Vol. 20, pp. 46-61, 1973

[26] F. Lo Presti, Z.-L. Zhang, D, Towsley, J. Kurose, "Source time scale and optimal buffer/bandwidth trade-off for regulated traffic in an ATM node," *Proceedings IEEE Infocom'97*, Kobe, Japan, April 1997.

[27] H. Sariowan, "A Service-Curve Approach to Performance Guarantees in Integrated-Service Networks," Ph.D. Dissertation, University of California, San Diego, 1996.

[28] H. Sariowan, R.L. Cruz and G.C. Polyzos, "SCED: a generalized scheduling policy for guaranteeing quality-of-service," submitted to *IEEE/ACM Transactions on Networking.*

[29] A. Skorokhod. 'Stochastic equations for diffusion processes in a bounded region', *Theory of Probability and its Applications*, Vol. 6, pp. 264–274, 1961.